SEMENT

INSTITUTE OF TECHNOLOGY

HD28
.M414
no1764-86

Dewey
APR 1 1987
LIBRARIES

WORKING PAPER
ALFRED P. SLOAN SCHOOL OF MANAGEMENT

MULTIPRODUCT QUEUEING NETWORKS WITH DETERMINISTIC ROUTING: DECOMPOSITION APPROACH AND THE NOTION OF INTERFERENCE

by
Gabriel R. Bitran*
and
Devanath Tirupati*

WP # 1764-86 March 1986

MASSACHUSETTS
INSTITUTE OF TECHNOLOGY
50 MEMORIAL DRIVE
CAMBRIDGE, MASSACHUSETTS 02139

MULTIPRODUCT QUEUEING NETWORKS WITH DETERMINISTIC ROUTING: DECOMPOSITION APPROACH AND THE NOTION OF INTERFERENCE

by
Gabriel R. Bitran*
and
Devanath Tirupati*

WP # 1764-86 March 1986

* Massachusetts Institute of Technology

M.I.T. LIBRARIES
APR 16 1987
RECEIVED

Multiproduct Queueing Networks with Deterministic Routing:

Decomposition Approach and the Notion of Interference

Gabriel R. Bitran and Devanath Tirupati

Massachusetts Institute of Technology, Cambridge, Mass.

ABSTRACT

Queueing networks have been used to model the performance of a variety of complex systems, however, exact results exist for a limited class of networks. The methodology that has been used extensively to analyse networks of queues is the decomposition approach. In this paper, we consider open queueing networks with multiple product classes, deterministic routings and general arrival and service distributions. We examine the decomposition method for such systems and show that it provides estimates of key parameters with an accuracy that is not acceptable in many practical settings. Recognizing this weakness, we enriched the approach by modelling a phenomenon previously ignored. We consider interference among products and describe its effect on the variance of the departure streams. The recognition of this effect has significantly improved the performance of this methodology. We provide extensive experimental results based on the data of a manufacturer of semiconductor devices.

1.0 Introduction:

Queueing networks have been used to model the performance of a variety of complex systems such as production job shops, flexible manufacturing systems, computers, urban service systems, communication networks etc. However, exact results exist for the limited class of Jackson type networks (Jackson 1957,1963 and Kelly 1975). For more general networks several approaches leading to approximate solutions have been proposed in the literature. Reiser and Kobayashi (1974) and Kuehn (1976, 1979) were among the first proponents of the parametric decomposition approach, which was later used by Shantikumar and Buzacott (1981) for single product networks and by Whitt (1983a,b). This approach generalizes the notion of independence and product form solutions of Jackson type networks to more general models (see section 2). In this method, the squared coefficients of variation (scvs) of interarrival intervals at each station are computed approximately. The performance measures such as mean number of jobs, queue lengths at each station are estimated based on these scvs.

In this paper, we consider an open queueing network with multiple product classes, deterministic routing for jobs in each product class and general arrival and service distributions. We examine the parametric decomposition approach for such a class and show that the approximations do not perform well in certain instances with large number of products. Our analysis shows that, under the assumption that the interdeparture intervals are independent and identically distributed (iid), the squared coefficient of variation (scv) of interdeparture intervals for each product stream leaving a station can be expressed as the sum of two terms. The first term reflects the influence of congestion and the service at the station while the second term represents the effect of the presence of other products in the network. This result can be interpreted as a generalization of the decomposition approach to a multiproduct network with deterministic routing. The generalization comes from the fact that we recognize the interference among products ie. at a given station, the scv of the departures of a product stream is distorted by the presence of other products. This is reflected in the second term mentioned above. In order to compute this interference effect for a product at a given station, we aggregate all other products into a single product. Therefore, the two product model is an important building block in our methodology.

Since it is difficult to determine the interference effect of the other products exactly, we obtain two approximations that are based on the following assumptions about the behavior of the aggregate product.

i) The arrivals of the aggregate product follow a Poisson process. This assumption is inspired by the results of Franken (1963) and Cinlar (1972) and relies on the notion that superposition of a large number of independent renewal processes can be approximated by a Poisson process.

ii) The interarrival intervals of the aggregate product have an Erlang distribution. This approximation is intended for cases with moderate number of products with an scv of less than .5 for interarrival intervals.

Both approximations are easy to compute and are asymptotically exact when the proportion of the demand due to each product tends to zero. Our computational results demonstrate that these are quite robust and provide reasonably accurate estimates for other distributions with the same scv.

The paper is organized as follows. In section 2 we provide a review of the literature and briefly describe the parametric decomposition approach. In section 3 we develop an approximation based on (i) above to compute the scvs for departure streams for networks with multiple products and deterministic routing. Computational experiments demonstrating the improvement due to the new approximation are described in section 4. The detailed study of the Erlang approximation and the related computational results are discussed in sections 5 and 6 respectively. In section 7 we examine these approximations in a network context. The network chosen is based on real life data and models the production facility of a semiconductor company. Finally, we summarize the results in section 8.

2.0 Review of the Literature:

The research on network of queues has focussed primarily on performance evaluation. For the purpose of a brief review of the literature, we classify the research as follows:

i) Exact Analysis.

ii) Approximation Methods.

iii) Simulation and related techniques.

2.1 Exact Analysis : Exact results exist for Markovian systems. A seminal contribution in this area is the paper by Jackson (1963) which provides the results for equilibrium probability distribution of the number of jobs for a variety of systems that are referred to as Jacksonian networks. This restricted class of networks can be described in the following framework.

1) The system consists of N stations or machine centers. Jobs which need processing at any station form a queue at that stage.

2) A first come first serve discipline is observed at each station.

3) The service time at each station follows an exponential distribution with parameter $\mu(i,k_i)$ where k_i is the total number of jobs at station i (in queue and in service).

4) The system state can be described by an N-dimensional vector $(k_1,k_2,...,k_N)$ where k_i is the number of jobs at station i, $i = 1,2,...,N$

5) The arrival process is Poisson with parameter $\lambda(K)$ where K is the total number of jobs in the system.

6) Job route is modeled as a random walk i.e. the next step on the job route depends only on the current station and not on the past processing history. Thus, the routing behavior can be described by a matrix R.

$R = \{r_{i,j}, i \in [0,N], j \in [1,N+1]\}$

where $r_{0,i}$ is the probability that an arriving job visits station i first.

$r_{i,j}$ is the probability that a job visits station j after being processed at station i, $1 \leqq i,j \leqq N$

and $r_{i,N+1}$ is the probabilty that a job leaves the system from station i.

The routing behavior is quite general enough to model arbitrary networks. Jackson provides sufficient conditions for the existence of an equilibrium distribution. The main result is that the equilibrium distribution, if it exists, is of a product form i.e.

$\pi(k)$ = Probability that the system is in state k in equilibrium

$= B\,\Gamma(k)\,\theta_1(k_1)\,\theta_2(k_2)...\theta_N(k_N)$

where k = state vector $(k_1,k_2,...k_N)$

$\Gamma(.)$ = a function of the number of jobs in the system.

$\theta_i(.)$ = a function that depends on the nature of station i. θ_i is proportional to the equilibrium probability distribution at station i assuming the arrivals at the station follow a Poisson process.

and B = a normalizing constant.

This result implies that to compute the equlibrium probability of a given state k, we can consider each station individually. Kelly (1975) extended Jackson's result to networks with more general routings and priority behavior. Specifically he considered a multiple product network with the following characteristics.

1) Job arrivals follow independent Poisson processes.

2) The service time at each station is homogeneous and has an exponential distributicn. The service time at a given station is independent of product type.

3) The routing behavior for each class can be described by a routing matrix as in a Jackson network. This permits deterministic routing for some or all product classes.

4) It is possible to define a variety of queue disciplines. $\beta_j(i,n_j)$ is the proportion of effort expended on job in position i at station j with n_j jobs. The service rate for this job would then be $\mu_j\ \beta_j(i,n_j)$. Ofcourse, the β_js together sum to one.

5) $\delta(i,n_j)$ is a function which defines another type of priority. $\delta(i,n_j)$ is the probability that a new arrival at station j with n_j jobs would occupy position i in the queue. Again, the δ's sum to one.

6) The state of the system is described by specifying the number of jobs at each station and the product type in each position of the queue.

Kelly showed that even in this more general network, the equilibrium probability distribution is of the product form i.e.

$\pi(k) = B.\ \Gamma(k).A_1(k_1).\ A_2(k_2)...\ A_N(k_N)$

where k = state vector,

k_i = a vector describing the product type of jobs at station i,

$\Gamma(.)$ = A function that depends on the state vector.

$A_i(.)$ = a function that depends on station i,

and B = a normalizing constant.

The product form result holds in the case of random order service discipline. When the service time is not homogeneous, the result holds with a preemptive resume, last come first serve discipline. In particular it does not hold for first come first serve policy and non-homogeneous service times. Similar results are provided for closed and mixed networks by Baskett et al (1975), Gordon and Newell (1967). The reader is referred to survey papers by Lemoine (1977) and Disney and Konig (1985) for more references and details of this approach.

While the product form results are interesting and useful, they are difficult to implement in practice due to the very large state space. Also, in many cases, the parameters of interest are the mean values of queue lengths, waiting times etc. and not $\pi(k)$. Reiser and Lavenberg (1980) consider closed networks for which the product form results hold. For such networks, they have developed a procedure (mean value analysis) to compute mean values without evaluating the state probabilities $\pi(k)$. The main result is a recursive relationship describing the performance measures of the system with K jobs as a function of those with K-1 jobs.

In summary, the exact results encountered in the literature basically rely on the following assumptions:

1) Exponential distribution for service times.

2) Homogeneous service requirement at all stations.

3) Priority discipline independent of customer class.

4) Poisson arrivals.

However, these are overly restrictive in many practical situations. Since the exact results do not extend to more general networks, it has led to the development of approximation schemes described in the subsequent sections.

2.2 Approximate Methods: The lack of success in obtaining exact solutions for general networks has motivated research in developing approximations to evaluate performance measures . These may be broadly described under the following four categories.

i) Diffusion approximation.

ii) Mean value analysis.

iii) Operational analysis.

iv) Decomposition methods.

The first three approaches are not directly related to the work in this paper and are mentioned here for completeness. We do not describe these in any detail, but provide references for the interested reader.

Diffusion Approximations are motivated by heavy traffic limit theorems and are based on asymptotic method for approximating point processes. The work by Iglehart and Whitt (1970), Newell (1971), Gaver and Shedler (1973), Kobayashi (1974), Gelenbe (1975), Harrison and Reiman (1981) is reperesentative of this type of analysis.

Mean Value Anlysis is a heuristic approach similar to the work by Reiser and Lavenberg (1980) and is intended for closed networks. Schweitzer (1979) proposed an approximation that has been extensively tested by Bard (1979). The resulting Schweitzer-Bard algorithm requires solution of a non-linear system to estimate the performance of closed network.

Buzen (1976) was among the first to use operational analysis computer systems. This approach focuses on directly measurable quantities and testable assumptions. The analysis is distribution free and relies on flow balance and homogeneous service principles. Denning and Buzen (1978) provide a tutorial on the approach. This paper contains a detailed list of references on the subject.

Decomposition Methods are essentially attempts to generalize the notion of independence and product form results for Jackson type networks to more general systems. This approach relies on two notions - a) the nodes can be treated as being stochastically independent and b) two parameter (typically mean and variance) approximations provide resonably accurate results. The approach was first proposed by Reiser and Kobayashi (1974) and has been used by several researchers in developing approximations. Sevick et al. (1977), Chandy and Sauer (1978), Kuehn (1979), Shantikumar and Buzacott (1981), Buzacott and Shantikumar (1985) and Whitt (1983a) have all used a similar approach. Essentially the method involves three steps-

Step 1: Analysis of interaction between stations.

Step 2: Decomposition of the network into subsystems of individual stations and their analysis.

Step 3: Recomposition of the results to obtain the network performance.

Step 1 is critical to the decomposition approach and we describe it in detail using the models of Shantikumar and Buzacott (1981) and Whitt (1983a,b). These are representative of the method and illustrate the approach. Also, we have used these as a bench mark to compare the approximations proposed in this paper. The interaction between the stations is analysed by looking at the network as a composite of three basic processes -

a) superposition or merging,

b) flow through a queue or a station, and

c) splitting or decomposition.

The first process typically represents the arrivals at a station while the third describes the departures from a station. The splitting and the merging processes model the product routing in the network. We first describe the approximations used by Shantikumar and Buzacott for the three basic processes. They considered networks with a single product, Poisson arrivals and general service times.

a) The merging process considers superposition of arrivals at a station (say station j). Let λ_{ij} and ca^i_j be the arrival rate and the scv of interarrival time of the flow from station i to j (i = 0 for external arrivals). The resulting mixture is approximately described with an arrival rate λ_j and scv ca_j as follows:

$$\lambda_j = \sum_{i=0}^{N} \lambda_{ij}$$

$$ca_j = \sum_{i=0}^{N} (\lambda_{ij}/\lambda_j)\, ca^i_j \quad (1)$$

b) The flow process describes approximately the scv of the departure stream from a station. If ca_j and cs_j are the scv of the interarrival and service time at station j with a utilization ρ_j, the scv of the departures is approximately given by

$$cd_j = \rho_j^2\, cs_j + (1-\rho_j^2)\, ca_j \quad (2)$$

c) In the decomposition process a product stream with arrival rate λ_j and scv of cd_j is split into a number of substreams, each stream representing the flow from a station j to other stations in the network. The routing is assumed to be Markovian. If p_i is the probability that a job would follow path i, then the substream i is characterized by the following parameters:

$$\lambda_{j,i} = p_i\, \lambda_j$$

$$cd_{j,i} = p_i\, cd_j + 1 - p_i \qquad (3)$$

Expression (3) implicitly assumes that interddeparture intervals of the product stream are iid with mean $1/\lambda$ and scv of cd. Note that $ca_i = cd_{j,i}$.

The interactions between the stations are reflected by the scv of the arrivals at each station. Shantikumar and Buzacott consider single product networks with Poisson arrivals and Markovian routing and general service time distribution. Combining the approximations for the three basic processes described above leads to the following linear systems.

$$a_i = \lambda r_{0,i} + \sum_{j=1}^{N} a_j r_{j,i}\,, i=1,2,...N \quad (4)$$

$$a_i ca_i - \sum_{j=1}^{N} \left\{ a_j (1-\rho_j^2) r_{j,i}^2 ca_j \right\} = \lambda r_{0,i} + \sum_{j=1}^{N} \left\{ a_j r_{j,i} (\rho_j^2 r_{j,i} cs_j + 1 - r_{j,i}) \right\}, i=1,2,...N \quad (5)$$

where λ = arrival rate into the system

a_j = net arrival rate at station j

ρ_j = utilization at station j

ca_j = scv of arrivals at station j

cs_j = scv of service at station j

$R = \{r_{i,j}\}$ = routing matrix

Note that (4) is the standard flow balance equations to determine the net arrivals at each station. The internal flows $\lambda_{i,j}$ are given by $a_i.r_{i,j}$. (5) is the approximation to compute the coefficient of arrivals at each station.

In step 2, each station is analysed based on the partial information obtained in step 1. Shantikumar and Buzacott examine M/G/1 and G/G/1 approximations to evaluate the station performance. In particular, the G/G/1 approximation requires only the mean and scv of the interarrival and service times.

Finally in step 3 these results are synthesized and performance measures for the network are estimated. This procedure is straight forward for mean queue lengths, and lead times.

In developing the queueing network analyser (QNA) Whitt (1983a) has essentially used an approach similar to the one described above in analysing the interactions between the nodes. The superposition or the merging process has been modified as follows:

$$ch_j = w\, ca_j + 1 - w \qquad (1a)$$

where ch_j is the modified scv of the merged stream, ca_j is determined by (1) and w is a weighting function that depends on the station utilization ρ_j and $\lambda_{i,j}$. In the QNA the weight w is determined as follows.

$$w^{-1} = \left[1 + 4(1-\rho_j)^2(v-1)\right]$$

$$where \quad v^{-1} = \sum_{k=1}^{m} \lambda_{kj}^2 / \lambda_j^2$$

This modification is motivated by the observation that neither the asymptotic method (which leads to 1) nor the stationary interval approach (see Whitt 1982 for details) by itself perform well over wide range of ca for approximating the superposition process. (1a) above is a hybrid approximation of both approaches and is based on the work of Whitt (1982) and Albin (1981,1982). Whitt also refines the approximation of the queue process to provide for multiple servers and uses the same approximation (3) for the splitting process. Estimation of station performance measures in step 2 are also modified to provide for multiple servers at each station. However, the approximations continue to use the first two moments of the arrival and service distributions.

These approximations are used in the QNA to analyse networks with multiple products and deterministic routing for each product. First, an aggregate product is formed by appropriately combining the product data and the analysis is done with the aggregate data. The aggregation procedure is briefly described below.

Let m = number of products

λ_k = arrival rate of product k

n_k = number of operations for product k

$n_{k,i}$ = station visited by product k at step i

1H(x) = 1 if x ε H and 0 otherwise.

The external arrival rates and the flow rates within the network are obtained as follows.

$$\lambda_{0j} = \sum_{k=1}^{m} \lambda_k 1\{k: n_{k,1} = j\}$$

$$\lambda_{ij} = \sum_{k=1}^{m} \sum_{l=1}^{n_k} \lambda_k 1\{(k,l): n_{k,l} = i,\ n_{k,l+1} = j\}$$

The routing matrix r_{ij} is defined by

$$r_{ij} = \frac{\lambda_{ij}}{(\lambda_{0,i} + \sum_{k=1}^{N} \lambda_{k,i})}$$

The mean and the scv of the service time at each station is computed by

$$\tau_j = \frac{\sum_{k=1}^{m}\sum_{l=1}^{n_k} \lambda_k \tau_{k,l} 1\{(k,l): n_{k,l}=j\}}{\sum_{k=1}^{m}\sum_{l=1}^{n_k} \lambda_k 1\{(k,l): n_{k,l}=j\}}$$

$$\tau_j^2(cs_j+1) = \frac{\sum_{k=1}^{m}\sum_{l=1}^{n_k} \lambda_k \tau_{k,l}^2(cs_{k,l}+1) 1\{(k,l): n_{k,l}=j\}}{\sum_{k=1}^{m}\sum_{l=1}^{n_k} \lambda_k 1\{(k,l): n_{k,l}=j\}}$$

where $\tau_{j,k}$ and $cs_{j,k}$ are the mean and scv of the service time for product j at step k.

Both Shantikumar and Buzacott and Whitt report encouraging results based on experiments comparing their respective approximations with simulation. However, the experiments are all (except for one case) based on a single product. The one exception is the case with two products presented in Whitt (1983b). However, in the manufacturing context, the problem with more than two products is more common and relevant and provides the motivation for the analysis in this paper.

2.3 Simulation Methods: In the absence of (exact) analytic results except for very restricted cases, discrete event monte-carlo simulation has been an obvious alternative to evaluate large queueing networks. This approach permits use of more elaborate assumptions that are closer to reality. The main drawback is the computational requirement. The process is time consuming and except in very small examples only a limited number of alternatives can be examined. Recent developments in perturbation analysis suggest considerable promise in reducing the computational needs. This technique, developed by Ho et al. (1979) provides a method to estimate, in one single simulation run, derivatives of performance measures with respect to the decision parameters. The evidence supporting the technique is to some extent experimental. Recent work by Suri and Zazanis (1984) and Zazanis and Suri (1984) examine some related theoretical issues.

3.0 SCV of Departures from a Station with Multiple Products-

Characterization of Interference in the presence of multiple products

The basis of the decomposition approach described in the previous section is the determination of mean and scv of arrivals at each station. It may be noted that specification of product arrival rates and routings determine the means exactly and hence the quality of the results depend on the scvs. In the case of multiple product networks with deterministic routing, this reduces to the estimation of scvs of each product stream. We observe that if the arrivals do not follow a Poisson distribution, the use of splitting process (3) to describe the scv of each product may not perform well in some instances (this conjecture is supported by the computations described in sections

4 and 6). In this section, we characterize, for each product, the scvs of departures from a single server station processing multiple products. In the process, we identify the distortion in the scv of a given product due to the presence of other products. We refer to this distortion as interference effect. The discussion in this section will make it clear that this interference effect may be estimated by aggregating all other products. Thus, in our methodology, the determination of interference reduces to the analysis of the two product case - product of interest and the aggregation of the other products. Our analysis is described below in detail and relies on the following assumptions.

i) The priority discipline at the station is FCFS.

ii) The interdeparture intervals from the station are iid.

iii) The product arrivals are independent. The interarrival times for each product are iid.

These assumptions are consistent with the decomposition approach. (ii) is clearly an approximation. (i) together with the single server assumption implies that the sequence of job arrivals (by product type) is identical to that of departures.

In this section, we consider a single server station processing multiple products and we characterize, approximately, the scv of departures for each product. The result is analogous to and generalizes the approximation for the splitting process given by (3). The interactions between stations in a network are now described by the approximations for the superposition (1) and queue (2) processes and this generalization.

Notation:

m = number of products

ρ = station utilization

cs = scv of service time

λ_i = arrival rate of product i

λ = arrival rate at the station = $\Sigma \lambda_i$

ca_i = scv of interarrival time of product i

x_i = interarrival time of product i (random variable)

$p_i = \lambda_i / \lambda$

d_i = interdeparture interval for product i (random variable)

cd_i = scv of interdeparture intervals of product i

d = inter departure interval from the station (random variable)

cd = scv of interdeparture intervals from the station

$n1_i$ = number of jobs of the aggregate product that arrive during an interarrival time of product i

$n_i = n1_i + 1$

cn_i = scv of n_i

$E(.)$ = expected value of (.)

$V(.)$ = variance of (.)

Note that the random variable d_i is the sum of n_i random variables which are iid by assumption (ii).

Hence, $E(d_i) = E(n_i)\,E(d)$

$= E(d)/p_i \quad = 1/(p_i\lambda) = 1/\lambda_i$

and $V(d_i) = E(n_i)\,V(d) + V(n_i)\,(E(d))^2$

since, $cd_i = V(d_i)/(E(d_i))^2$

$= V(d)/(E(n_i)\,(E(d))^2) + V(n_i)/(E(n_i))^2$

we get, $cd_i = p_i\,cd + cn_i \qquad (6)$

The scv of the departures of each product stream is characterized approximately by (6). It is interesting to note that (6) expresses the scv of the departure stream as the sum of two terms. The first term can be considered as reflecting the effect of the queue process while the second term captures the interference due to other products. This is in contrast to that given by (3) which does not explicitly recognise the latter effect. The expression in (6) may be interpreted as a generalization of (3) for multiple products. It reduces to (3) if either a) the routing is Markovian with probability p_i or b) the arrivals at the station are Poisson ($ca = 1$). Also, note that, in the case of single product and deterministic routing (6) reduces to cd.

We observe that (6) is not easy to implement, since it is difficult to evaluate cn_i in the general case. In the remainder of the paper, we propose and test approximations to determine cn_i. These approximations are based on assumptions about the behavior of the aggregate product.

3.1 Poisson approximation for the aggregate product

In this approximation we assume that the arrivals of the aggregate product follow a Poisson process. This is motivated by the notion that superposition of a large number of independent renewal processes may be approximated by a Poisson process and is inspired by the results of Franken (1963) and Cinlar (1972). In what follows, we derive expressions for cn_i and cd_i. For simplicity, we omit the subscript i from X_i, the interarrival time for product i.

Let $f_X(x)$ be the probability density function of the interarrival interval for product i,

and let $E(X) = 1/\lambda_i$

The arrivals of products other than i follow a Poisson process with parameter $\lambda(1\text{-}p_i)$

First, we derive the probability distribution of $n1_i$

$Pr[n1_i = n]$ = probability that $n1_i$ takes value n

$$Pr[n1_i = n] = \int_0^\infty Pr[n1_i = n | X = x] f_X(x)\,dx \ ,$$

$$since \qquad Pr[n1_i = n | X = x] = \frac{(\lambda(1-p_i)x)^n}{n!} exp(-\lambda(1-p_i)x)$$

$$we\ have \qquad Pr[n1_i = n] = \int_0^\infty \frac{(\lambda(1-p_i)x)^n}{n!} exp(-\lambda(1-p_i)x) f_X(x)dx$$

It can be shown that

$$E(n1_i) = \sum_{n=0}^{\infty} nPr[n1_i = n] = (1-p_i)/p_i$$

$$and \qquad E(n_i) = 1/p_i$$

Next, we compute $E(n1_i^2)$ as follows:

$$E(n1_i^2) = \sum_{n=0}^{\infty} n^2 Pr[n1_i = n] = \sum_{n=0}^{\infty} n(n-1)Pr[n1_i = n] + \sum_{n=0}^{\infty} nPr[n1_i = n]$$

$$and \quad \sum_{n=0}^{\infty} n(n-1)Pr[n1_i = n] = \int_0^\infty \sum_{n=2}^{\infty} \frac{(\lambda(1-p_i)x)^{n-2}}{(n-2)!} (\lambda x(1-p_i))^2 e^{-\lambda x(1-p_i)} f_X(x)dx$$

$$it\ can\ be\ shown\ that \quad \sum_{n=0}^{\infty} n(n-1)Pr[n1_i = n] = (1+ca_i)(1-p_i)^2/p_i^2$$

$$hence \quad E(n1_i^2) = \frac{(1-p_i)}{p_i} + (1+ca_i)\left[\frac{(1-p_i)}{p_i}\right]^2$$

$$and\ V(n1_i) = E(n1_i^2) - (E(n1_i))^2 = \frac{(1-p_i)}{p_i} + ca_i\left[\frac{(1-p_i)}{p_i}\right]^2.$$

$$Since\ V(n_i) = V(n1_i), \quad cn_i = p_i^2 V(n_i) = (1-p_i)\left[p_i + (1-p_i)ca_i\right]$$

An approximation for cd_i is given by

$$cd_i = p_i cd + (1-p_i)\left[p_i + (1-p_i)\ ca_i\right] \quad (7)$$

Note that approximation (7) reduces to (3) when the arrivals are Poisson (or $ca_i = 1$) and thus it can be considered to be a generalization of the approximations used by Whitt and Buzacott and Shantikumar. However, it differs from (3) even qualitatively in the general case. For example, as p_i approaches zero, (3) would suggest that cd_i tends to one, where (7) indicates that cd_i approaches ca_i. As the following proposition demonstrates, (7) is exact in the limit.

Proposition: Consider a single server station with multiple products, arrival rate λ, utilization ρ and a given service distribution. Assume that the arrivals to the station can be approximated by iid interarrival times. Then as $p_i \to 0$, $cd_i \to ca_i$ for all i.

Outline of Proof: Let w_j be the waiting (queue + service) time for job j. Note that, in equilibrium, the w_js are identically distributed for all j. Let w be the equlibrium waiting time. The station can be modelled as a G/G/1 queue. An upper bound on V(w), the variance of w is given by $\sigma_a^2 + 2\,\sigma_b^2$, where σ_a and σ_b are the standard deviations of the interarrival and service times respectively. Let V(w)* be the upper bound on V(w). (For details see Kleinrock, 1976).

Consider an interdeparture interval for product i. Let x_i be an interarrival interval for product i and w_1 and w_2 be the corresponding waiting times for the two successive jobs of this product.

$$\textit{Then,}\quad d_i = x_i - w_1 + w_2 \qquad \textit{and} \quad V(d_i) = V(x_i) + V(w_1 - w_2) - 2\,Cov(x_i, w_1 - w_2)$$

$$\textit{Since}\ V(w_1 - w_2) \le 4\,V(w) \le 4\,V(w)^* \ \textit{and}\ |Cov(x_i, w_1 - w_2)| \le [V(x_i)\,V(w_1 - w_2)]^{0.5} \le 2[V(x_i)V(w)^*]^{0.5}$$

$$\textit{we have,}\ V(x_i) - 4\,[V(x_i)\,V(w)^*]^{0.5} \le V(d_i) \le V(x_i) + 4\,V(w)^* + 4\,[V(x_i)\,V(w)^*]^{0.5}$$

Note that V(w)* does not depend on p_i. Also note that in equilibrium, $E(d_i) = E(x_i) = 1/\lambda_i = 1/(\lambda p_i)$. Hence,

$$ca_i - 4\,\lambda\,[ca_i\,V(w)^*]^{0.5}\,p_i \le cd_i \le ca_i + 4\,\lambda^2\,V(w)^*\,p_i^2 + 4\,\lambda\,[ca_i\,V(w)^*]^{0.5}\,p_i$$

Since λ and V(w)* are constant, as $p_i \to 0$, $cd_i \to ca_i$.

As the number of products increases, we can expect p_i to be small and thus the approximation (7) is in some sense asymptotically exact in the number of products. The formula has other implications as well for queueing networks. If the number of products processed at each station is large, we can ignore the interaction between stations and analyse each station independently. The mean and variance of each product stream would be preserved throughout the network and we could assume that at every station visited by a product, the first two moments would be the same as those at the time of the external arrivals into the system.

We make two remarks about the quality of approximation (7). First, we can expect that the approximation will perform well as the arrivals are close to Poisson and the performance will deteriorate with increasing deviations from the Poisson process. Second, we can expect the approximation will be good for small values of p_i (and large number of products). In the following section we describe results of computations to examine the goodness of this approximation.

4.0 Computational Results

In this section we report the results of computational experiments comparing the various approximations. In the absence of any exact results, we use simulation as a bench mark to evaluate the approximations. This approach is common enough in the analysis of queueing networks. For

example see Fraker (1971), Kuehn (1979), Shimshak (1979), Shantikumar and Buzacott (1981), Whitt (1983b, 1984) etc. The experiments were designed to examine the scv of the departure streams from a single station processing multiple products. We considered the following four factors in the design of the test cases.

1) Number of products (and the fraction p): Four levels were considered for this factor. Number of products were set at 2,3,5 and 10. In each case the arrival rates and distributions of interarrival times were identical for all the products. The fraction of demand due to each product (p) ranged from 0.1 to 0.5.

2) The distribution of interarrival time: These were assumed to be iid with Erlang distribution. The Erlang parameter was set at 2,3 and 4 to yield 3 levels for this factor.

3) Station Utilization: Two levels of station congestion, $\rho = 0.6$ and 0.9 were tested. These were considered representative of moderate and high utilizations in manufacturing settings.

4) Service time distribution: The service times were assumed to be iid with Erlang distribution. The parameter of the distribution was set at 2 and 3.

Thus, in all, 48 problems were simulated. The choice of Erlang distribution was motivated by the fact that in many manufacturing environments the variability in process times and job releases is much smaller than that of a Poisson process i.e. scvs are typically smaller than one. The Erlang family provides a range of scvs in this domain. It may be noted that the Erlang distribution has been used in the reported literature to examine cases with scvs less than one. In addition, we consider some other distributions in the experiments reported in section 6.

To evaluate the performance of the approximations, we compare the results with the estimates provided by the following alternatives.

1) Approximation (7) together with (1) and (2). This approach, which uses the notion of interference introduced in this paper is referred to as INT1 in subsequent discussions.

2) Simulation.

3) Application of approximations (1a), (2) and (3) to the aggregate product defined by Whitt (1983a) and described earlier. The routing matrix for the aggregate product was given in section 2. We refer to this as the aggregation approach and the approximation is denoted by AGP1 in the rest of the paper.

4) Application of approximations (1), (2) and (3) for the aggregate product which is defined as in AGP1. This is a generalization of the approach by Shantikumar and Buzacott to networks with multiple products and we denote this approximation by AGP2.

Figures 1 and 2 describe the behavior of the approximations as a function of the fraction p for two data sets specified by ρ, ca and cs and are representative of the results obtained. The figures compare the estimates of cd_i and clearly bring out the fact that INT1 converges to that given by the simulation as p tends to zero. In contrast, the approximations AGP1 and AGP2 which are based on

random routing provide estimates that approach the value one. It may be observed that INT1 represents substantial improvement over both AGP1 and AGP2 in the estimates for cd_i. Note that INT1 requires an estimate for scv of station departures cd. We use (1) and (2) to estimate this parameter. The same procedure is used for AGP2 while (1a) and (2) are used for AGP1. The differences in the estimates of cd_i arise from the use of (3) and (7) for the splitting process.

Tables 1-4 describe the results in detail and compare the estimates of cd_i for the three approximations with the simulation value. In each case, by symmetry, the cd_i values are the same for all the products and the simulation value reported in the tables represent the mean value obtained in the experiments. These results support the analysis of the previous section and permit the following conclusions.

1) The computational experiments suggest that INT1 is a distinct improvement over AGP1 and AGP2 in the cases tested. For example, consider the ten product case when ca is 0.333 in table 4. The AGP1 and AGP2 estimates are .976 and .935 compared to the simulation value of .36. The INT1 estimate is .39 which is within 10%. The results are less dramatic but still significant for the two product case - 0.81 and .75 by AGP1 and AGP2 and .625 of INT1 compared to .545 given by simulation.

2) The approximations deteriorate as the arrivals diverge from the Poisson process. For example, consider the ten product case in table 2. The error is less than 6% when ca is .5 (estimate of .545 for .5). The errors become progressively worse as the ca value decreases - error of 10% when ca is .333 and 15% when ca takes the value .25.

3) As expected the quality of the approximation improves with the number of products (smaller p_i).

4) The approximations over estimate the value of cd_i.

The above conclusions should be tempered by the fact that the measures of interest are mean queue lengths, number of jobs, and waiting times at the stations and not the coefficient cd itself. Hence, it is necessary to examine the impact of errors in cd on the queue length estimates at the subsequent stations. Note that the cd_i of the departure stream represents the arrivals at subsequent stations and should be interpreted as ca at the next station. We observe that the two moment approximations to evaluate the queue lengths are of the form

$$L_q = \rho^2/(2(1-\rho))\,(ca+cs)g, \qquad L = \rho + L_q$$

where L_q = the mean queue length at the station

L = mean number of jobs at the station

and $g = f(ca,cs,\rho) = \exp[-2(1-\rho)(1-ca)^2/(3\rho(ca+cs))]$

This approximation is due to Kraemer and Langenbach-Belz (1976) and has been used by both Whitt and Shantikumar and Buzacott. The latter, in fact, have used two other approximations. The choice of the appropriate formula depends on the values of ca and cs. The function g is not very

sensitive to either ca or cs. Thus, if ca and cs are of comparable magnitude, the impact of an x% error in the estimate of ca would result in an error of magnitude (x/2)% in the estimate of L_q. The error in the estimate of L would be even lower. It is encouraging to note that large errors in the estimate of cd_i occur when the value is small and the impact on queue length estimates is unlikely to be large. When the cd_i are large, the errors are small and would not affect the queue length estimates substantially.

To illustrate this effect, we consider a six station, five product network shown in figure 3. All the products are first processed at station 1 and need a second operation at stations 2-6. Note that the departure streams from station 1 correspond to the experiments described earlier. We now examine the mean number of jobs at stations 2-6 for two cases. The two sets of parameters are described in table 5 and the results are given in table 6. Observe that, by symmetry, the mean number of jobs will be the same at stations 2-6 and the simulation result reported in table 6 represents the average of the values observed at these stations. The table provides, in addition to AGP1, AGP2 and INT1 estimates, those resulting from M/M/1 and M/G/1 approximations. The figures in the parenthesis in the table measure devations of the estimates from the simulation result. We also provide an estimate of the mean number of jobs computed from an error free estimate of the scv of the departures from station 1 by simulation. The intent is to illustrate the performance of the approximation to estimate the mean queue lengths.

The results in the table are encouraging. For case 1, the error in the mean number of jobs is 4.2% in contrast to the 12% error in the estimate of cd_i. The corresponding figures for case 2 are 9.7% and 18.2% respectively. These result provide additional evidence of the improvement by INT1 over AGP1 and AGP2. For example, in case 2 AGP2 and AGP1 overestimate the parameter by 52.3% and 57.9% respectively. The M/M/1 approximation performs poorly with an error of 124%. The reader should note that when the arrivals are Poisson ($ca_i = 1$ for all i), the computational results presented in Whitt (1984) indicate that AGP1 and AGP2 perform well. In that case INT1, AGP1 and AGP2 are the same.

5.0 Erlang Approximation

In this section we study, in detail, the scv of departure streams from a single server station processing two products. As explained earlier, the two products represent the product of interest (whose cn_i is to be determined) and the aggregation of all the other products. We propose two additional approximations INT2 and INT3 to estimate the interfernce effect. These are motivated by the fact that INT1 does not perform very well for small values of ca_i. The approximations proposed in this section are based on the assumption that the interarrival times of the two products have an Erlang distribution. The choice of the Erlang is due to the fact that it provides a family of distributions with scv smaller than one and this is of interest in many manufacturing settings. Further, this assumption facilitates the analysis and leads to computationally viable approximations. The computational results of the following section demonstrate that the scvs are not sensitive to the

specific distributions but depend on the first two moments only and hence the approximation can be used with other distributions also. In the analysis which follows, we assume that

i) Arrivals at the station belong to two product classes - product of interest and the aggregation of the other products.. The arrivals in each group are independent of each other.

ii) The interarrival times for each product are independent and have an Erlang distribution. The parameters of the distribution are λ_1, k_1 and λ_2, k_2 for products 1 and 2 respectively. (Product 2 is the aggregate product). Note that ca_i is $1/k_i$ for i = 1 and 2.

iii) As before, we assume that interdeparture intervals from the station are iid with a scv of cd.

$$let \quad p_i = \frac{\lambda_i/k_i}{\lambda_1/k_1 + \lambda_2/k_2} \ , \ i=1,2$$

An estimate of cd_i is given by (6) , $cd_i = p_i cd + cn_i$

We describe below,the two approximations INT2 and INT3 to estimate cn_i.

5.1 Approximation INT2 In evaluating cn_1 (cn_2), we assume that during any interarrival interval of product 1 (2), the arrivals of jobs of product 2 (1) follow an independent Erlang ditribution with parameters λ_2, k_2 (λ_1, k_1). We are, in effect ignoring the dependence in the arrivals of product 2 (1) between successive interarrival intervals of product 1 (2). However, note that, this is an approximation for the first arrival of product 2 (1) only and it is exact for subsequent arrivals. We now derive the probability distribution of $n1_1$ the number of jobs of product 2 that arrive during an interarrival time of product 1.

Let $f_{X1}(t)$ and $f_{X2}(t)$ be the probability density functions of interarrival intervals for products 1 and 2 respectively.

$$f_{X1}(t) = \lambda_1^{k_1} \frac{t^{k_1-1}}{(k_1-1)!} exp(-\lambda_1 t) \ , \ t \geqq 0$$

$$f_{X2}(t) = \lambda_2^{k_2} \frac{t^{k_2-1}}{(k_2-1)!} exp(-\lambda_2 t) \ , \ t \geqq 0$$

Property 1: Let X be a random variable with an Erlang distribution with parameters λ and k. The cumulative distribution of X is given by the following:

$$F_X(t) = 1 - \left[\sum_{i=0}^{k-1} \frac{(\lambda t)^i}{i!} \right] e^{-\lambda t} \ , \ t \geqq 0 .$$

Property 2 The probability density function for n arrivals (n-fold convolution) of product 2 is Erlang with parameters λ_2 and nk_2. Hence,

$$f^n_{X2}(t) = \lambda_2^{nk_2} \frac{t^{nk_2-1}}{(nk_2-1)!} exp(-\lambda_2 t) \ , \ t \geqq 0 .$$

The probability distribution of $n1_1$ is obtained as follows

$$Pr[n1_1 = n] = \int_0^\infty Pr[n1_1 = n | X1 = x] f_{X1}(x) dx$$

$$But, \quad Pr[n1_1 = n | X1 = x] = Pr[n1_1 \geqq n | X1 = x] - Pr[n1_1 \geqq n+1 | X1 = x]$$

$$and \quad Pr[n1_1 \geqq n | X1 = x] = \int_0^x f_{X2}^n(t) dt = 1 - \sum_{i=0}^{nk_2 - 1} \frac{(\lambda_2 x)^i}{i!} exp(-\lambda_2 x)$$

$$hence, \quad Pr[n1_1 = n | X1 = x] = \sum_{i=nk_2}^{(n+1)k_2 - 1} \frac{(\lambda_2 x)^i}{i!} exp(-\lambda_2 x)$$

$$and \quad Pr[n1_1 = n] = \int_0^\infty \sum_{i=nk_2}^{(n+1)k_2 - 1} \frac{(\lambda_2 x)^i}{i!} exp(-\lambda_2 x) \lambda_1^{k_1} \frac{x^{k_1 - 1}}{(k_1 - 1)!} exp(-\lambda_1 x) \, dx$$

$$= \sum_{i=nk_2}^{(n+1)k_2 - 1} \int_0^\infty \frac{\lambda_1^{k_1} \lambda_2^i}{i!(k_1 - 1)!} x^{i + k_1 - 1} exp\{-(\lambda_1 + \lambda_2)x\} \, dx$$

$$hence, \; Pr[n1_1 = n] = \frac{q^{k_1}}{(k_1 - 1)!} \left[\sum_{i=nk_2}^{(n+1)k_2 - 1} \frac{(i + k_1 - 1)!}{i!} (1 - q)^i \right], \; where \; q = \frac{\lambda_1}{\lambda_1 + \lambda_2} \quad (8)$$

It can be verified that the distribution obtained for $n1_1$ is a proper probability mass function. Clearly, $Pr[n1_1 = n]$ is nonnegative and it can be shown that these probabilities sum to one. It is not easy to obtain closed form expression for $V(n1_1)$, the variance of $n1_1$. But it is straight forward to compute it for given values of the parameters by using (8) directly. Since n_1 is $n1_1 + 1$, we have the following.

$E(n_1) = E(n1_1) + 1 \qquad$ and $V(n_1) = V(n1_1)$

$cn_1 = V(n1_1)/[E(n1_1) + 1]^2$

Hence cn_1 can be computed from the parameters of the arrival distributions. The coefficient for the second product can be computed in an analogous fashion.

5.2 Approximation INT3 : In this approximation we assume that the arrival of product 1 is an instance of random incidence in the arrival stream of product 2. The distribution of the time till the first arrival of product 2 during an interarrival interval of product 1 is then given by the following

$$f_{X2r}(x) = \frac{1}{k_2} \sum_{i=1}^{k_2} \lambda_2^i \frac{x^{i-1}}{i!} \exp(-\lambda_2 x) \quad , \; x \geq 0$$

The distribution for the subsequent arrivals is Erlang with parameters λ_2, k_2. Thus INT3 differs from INT2 in the assumptions regarding the distribution of time till the first arrival. The distribution of $n1_1$ can be derived as in the previous case. It can be shown that

$$Pr[n1_1 = 0] = \frac{q^{k_1}}{(k_1-1)!\,k_2} \left[\sum_{j=1}^{k_2} \sum_{i=0}^{j-1} \frac{(i+k_1-1)!}{i!} (1-q)^i \right]$$

$$\text{and} \quad Pr[n1_1 = n] = \frac{q^{k_1}}{(k_1-1)!\,k_2} \left[\sum_{j=1}^{k_2} \sum_{i=(n-1)k_2+j}^{nk_2+j-1} \frac{(i+k_1-1)!}{i!} (1-q)^i \right] \text{ for } n > 0 \quad (9)$$

Again, we do not have expressions in closed form for variance of $n1_1$ and n_1. For specified parameters, cn_1 can be computed directly from (9).

We make a few remarks about INT2 and INT3. First, the two approximations are very similar and for low values of p_i we do not expect any significant difference in the two estimates. Second, the approximation is only in respect of the first arrival. Hence, if the number of jobs arriving during an interarrival interval is large (or p is small), we may expect INT2 and INT3 to give good results. INT2 is likely to perform poorly when the number of jobs arriving during an interarrival time is very small. This will happen most often when the fraction p_i is close to one. We also expect the applicability of the two approximations to be more general than the case for which they were derived. This hope is based on the fact that two moment approximations provide reasonable results in most cases. Hence, we expect INT2 and INT3 to perform well for other arrival distributions with the same scv. We empirically examine some of these issues in the computational experiments described in the next section.

6.0 Computational experiments to test Erlang approximations:

In this section we empirically test the goodness of approximations INT2 and INT3 and report the results of the computational experiments. In the first set of experiments we compare the estimates of cd_i by INT2 and INT3 with other approximations and simulation results for a two product, one station system. In the design of these experiments the following factors were considered.

i) The fraction p_i is the proportion of demand at the station due to product i. In the experiments of section 4, p_i was the same for all the products (1/m). In the present case, we consider nine levels of p_i, 0.1 to 0.9 in steps of 0.1 to analyse impact of changing the product proportion.

ii) The arrival distribution: Job arrivals in each product class are independent of each other. The scvs of the interarrival times for each product are set at .333. We consider three distributions with the same scv - a) Erlang distribution with k=3, b) Uniform distribution over [0,a], and c)

Beta distribution with parameters 2 and 6. The intent is to examine the robustness of the approximations to distributions with the same scv.

iii) Station utilization: As in section 4, we consider two levels of congestion, $\rho = .6$ and .9.

iv) Service distribution is assumed to be Erlang with parameters 2 and 3.

In all 108 problem sets were simulated. Figures 4 and 5 describe the behavior of the approximations for two data sets defined by station utilization (ρ) and service distribution (cs). The simulation values in the figures correspond to the Erlang arrival case. The results with other distributions are not shown because they are very close to this value. The figures demonstrate the improvements in the cd estimates by INT1, INT2 and INT3 over AGP1 and AGP2. For low values of p, INT2 and INT3 give comparable results and appear to perform better than the others. On the otherhand, for larger values of p, INT3 seems to be the best method. This is consistent with our expectations. The detailed results are furnished in tables 7-10 and lead to the following conclusions.

i) INT2 and INT3 are very good approximations in the test cases. The estimates of cd are typically within 10% of the simulated values. While the performance of INT2 deteriorates a little with increasing p_i, INT3 consistently gives good estimates of cd.

ii) The scv of the departure streams is not sensitive to the arrival distributions. The range of the cds for the three distributions tested is less than 5% in most cases. This supports our conjecture that the approximations are robust and should give reasonable results for distributions with the same scvs.

We now examine the behavior of INT2 and INT3 for the multiproduct case. The test problems are those described in section 4. The cd_i estimates given by INT2 and INT3 are displayed in table 11. The table also provides the simulation values for comparison. From figures 1 and 2 and the tabulated results, it is obvious that INT2 and INT3 perform better than the others even for the multiproduct case. The value of p_i is not greater than 0.5 and both INT2 and INT3 give good results. These results suggest that with increasing number of products, INT2 performs better than INT3. This might be explained by the fact that as the number of products increases, the aggregate product departs from an Erlang process, and the corresponding scv of the interarrival time tends to increase. In such situations, the random incidence considered in INT3 becomes gradually counter productive in modelling the interference phenomenon. We would expect that as the number of products tends to infinity, INT1 will outperform both INT2 and INT3. The errors in the cd estimates are typically within 5% (the maximum error is 12%) which would suggest that the errors in the queue length estimates are likely to be very small.

Finally to see the impact of the Erlang approximation on the queue lengths, we consider the five product, six station example of section 4. The INT2 and INT3 estimates for mean number of jobs and the simulation results are displayed in table 12. The error in the INT2 estimate for the two cases

are 1.51 and 2.13% compared to 27.7 and 52.3% given by AGP2, and 30.67 and 57.9% given by AGP1. The corresponding figures for the INT3 estimates are 1.51 and 1.12% respectively.

7.0 A Network Example

To recapitulate, the intent of the approximations presented in this paper is to evaluate performance measures in a queueing network and in this section we report the results of an experiment in this context. The network for this set of experiments is based on real life data and models the production facility of a semiconductor company. The facility is represented by 13 machine stations and processes jobs of 10 product families. The routing for jobs in each family is deterministic. The network characteristics are described, in detail, in the appendix. Note that product families visit station 2 more than once. We considered the following factors in designing the problems for this experiment.

1) Arrival Distribution: The interarrival intervals for each product family were assumed to be iid. For each product the arrival distribution was randomly assigned from five alternatives - Erlang with parameters 2,3,4, exponential and uniform distribution.

2) The same five alternatives were considered for the distribution of service time. For each station, the distribution of service time was randomly assigned.

3) Station utilization was assumed to be uniformly distributed between 0.65 and 0.95. Again, the utilization at each station was determined in a random manner.

We generated ten problems for this experiment. Summary results for the ten problems are presented in table 13. In the table we compare the estimates of mean number of jobs in the network by approximations INT1, INT2 and INT3 with those given by AGP1, AGP2 and simulation. The results demonstrate the improvements that can be obtained by explicitly recognizing the interference term. Overall, INT2 seems to be the best procedure, while in most cases INT3 provides errors between those of INT1 and INT2. These results are consistent with those in tables 1-4 and 11. The average error with INT2 and INT1 are 1.78% and 5.95% respectively. In contrast, the average error with AGP2 and AGP1 are 23.67 and 27.39% respectively. The errors in the estimates by the aggregation approach (AGP1 and AGP2) are typically above19.45% in the problems tested. Case 10 is an exception with errors of 12.5 and 15.26%. This problem corresponds to the case in which the arrivals of five of the ten products follow a Poisson process. It is also not surprising that this is one of the two cases in which approximation INT1 (error 1.47%) performs better than INT2 (error 3.76%). The estimates at the station level are not as good but are still substantially better than those given by AGP1 and AGP2. In the ten test problems, the error in the estimate of mean number of jobs was within 10% at 112 stations (86% of the observations). The corresponding figures for the other approximations were 93 (71.5%), 14 (10.77%) and 9 (6.92%) for INT1, AGP2 and AGP1 respectively. We present the summary results for the station level statistics in table 14. It may be noted that, with the aggregation approach, the error is more than 20% at nearly 60% of the stations. The maximum

error is 64 and 63%for AGP1 and AGP2 respectively. In contrast, the maximum error with INT2 is 18.8% .

We make two remarks about the use of the Erlang approximations INT2 and INT3 in these estimates. First, note that for a given p_i, cn_i can be estimated for discrete values of arrival scvs. Since the scvs in the network are continous variables, we obtain cn_i by linear interpolation in such cases. Secondly, since the cn_is need to be consistent with these scvs, an iterative procedure is required to determine them. In the test problems, typically three to four iterations were required.

8.0 Conclusions:

In this paper we have examined the use of the parametric decomposition approach to the analysis of multiproduct queueing networks with deterministic routings. We have demonstrated that the approximation for the disaggregation or the splitting process based on Markovian routings can be quite poor. The analysis in this paper shows that the scv of product departures can be approximated as the sum of two terms. The first term represents the influence of station congestion and service time while the second term is the interference term due to the presence of other products. This interference term explicitly recognizes the presence of multiple products and can be considered as a generalization of the one used by Shantikumar and Buzacott, Whitt and others in the decomposition approach.

Since the estimation of the interference effect is difficult to evaluate exactly, we proposed and tested three approximations that are, in some sense, asymptotically exact. The first approximation (INT1) is based on the notion that the aggregation of the arrival streams of a large number of products may be approximated by a Poisson process. The computational results show that INT1 performs reasonably well over the range of parameters tested. The quality of the approximation deteriorates a little as the scv of arrivals diverge substantially from one. The other approximations, INT2 and INT3 are based on an Erlang distribution approximation for the arrivals and are intended for arrivals with scvs much smaller than one (for example, scvs smaller than 0.5). While INT2 and INT3 give comparable results for low values of p, INT3 gives better estimates at high values of p (greater than 0.5). However, INT2 seems to perform better than INT3 as the number of products increases. Our computational results show that approximations INT1, INT2 and INT3 give substantially superior results compared to the aggregation approach (AGP1 and AGP2) and the MG1 and MM1 approximations. This is demonstrated by the five product, six station example. The error in the estimate of the mean number of jobs is 2.13% by INT2 compared to errors in excess of 50% with the aggregation approach. The experiments also suggest that the results are robust with respect to the arrival distributions and the critical factor is the scv. This is in the same spirit as the two moment approximations in the decomposition approach

The potential of the approximations proposed in this paper is demonstrated by the network experiments with representative data from a semiconductor manufacturing company. The

application of INT2, INT3 and INT1 give estimates that are typically within 5 and 10% of the simulation figures. This is a substantial improvement over the approximations AGP1 and AGP2 that result in errors of over 20% in most cases.

In table 15 we summarize our recommendations regarding the application of INT1, INT2 and INT3. These are based on the computational results which support the preceeding analysis. We recommend the use of INT1 when the scv of arrivals is greater than .5. When the scv of arrivals is not greater than 0.5, INT2 or INT3 may be used for low values of p. For larger values of (greater than 0.5), we recommend use of INT3. Finally, we observe that multiple products and deterministic routings are representative of many manufacturing environments. Also, the variability in job arrivals and service times is, according to our experience, usually less than that of a Poisson process and scvs smaller than one are typical. This enhances the potential for the approximations proposed in this paper.

Acknowledgements:

The authors are grateful to Professors Hirofumi Matsuo and Stephen C. Graves for their helpful comments on an earlier version of this paper.

References

Albin, S.L., "Approximating queues with superposition arrival processes," PhD dissertation, Dept. of IE and OR, Columbia Univ., 1981.

Albin, S.L., "Poisson approximations for superposition arrival processes in queues," Mgt. Sc., 28, 2, 1982, 126-137.

Bard, Y., "Some extensions to multiclass queueing network analysis, Performance of Computer Systems," M.Arato, A.Butrimenko and E.Gelenbe (eds.), 1979, North Holland.

Baskett,F., K.M. Chandy, R.R. Muntz and F.G. Palacios, "Open, closed and mixed networks of queues with different classes of customers," J.ACM 22, 2, Apr 1975, 248-260.

Buzacott, J.A. and J.G. Shantikumar, "Approximate queueing models of dynamic job shops," Mgt. Sc., 31, 7, 1985, 870-887.

Buzen, J.P., "Fundamental operational laws of computer system performance," Acta Inf vol 7, no. 2, 1976, 167-182.

Chandy, K.M., and C.H. Sauer, "Approximate methods for analysing queueing network models of computer systems," ACM Computing Surveys, vol 10, no. 3, 1978, 281-317.

Cinlar, E., "Superposition of point processes," Stochastic Point Processes: Statistical Analysis, Theory and Applications, P.A. Lewis, ed., Wiley, NY, 1972, 549-606.

Denning, P.J. , and J.P. Buzen, "The operational analysis of queueing network models," ACM Computing Surv., 10, 3, sept 1978, 225-261.

Disney, R.L., and D. Konig, "Queueing networks: A survey of their random processes," SIAM Review, 27, 3, 1985, 335-403.

Fraker, J.R., "Approximate techniques for the analysis of tandem queueing systems," PhD dissertation, Dept. of I.E., Clemson University, 1971.

Franken, P., "A refinement of the limit theorem for the superposition of independent renewal processes," Theory Prob. Appl., vol 8, 1963, 322-334.

Gaver, D.P. and G.S. Shedler, "Approximate models for processor utilization in multi-programmed computer systems," SIAM J. Comput., 2, 1973, 183-192.

Gelenbe, E., "On approximate computer system model," JACM, 22, 1975, 261-269.

Gordon, W.J., and G.F. Newell, "Closed queueing systems with exponential servers," Oper. Res., 15, 2, Apr 1967, 254-265.

Harrison, J.M., and M.I. Reimann, "On the distribution of multidimensional reflected Brownian motion," SIAM J. App. Math., 41, 2, 1981, 345-361.

Ho, Y.C., M.A. Eyler, and T.T. Chien, "A gradient technique for general buffer storage design in a serial production line," Int. Jour. of Prod. Res, 17, 6, 1979, 557-580.

Iglehart, D.L., and W. Whitt, "Multiple channel queues in heavy traffic, II: Sequences, networks and batches," Adv. App. Prob., 2, 2, 1970, 355-369.

Jackson, J.R., "Networks of waiting lines," Oper. Res., 5, 1957, 518-521.

Jackson, J.R., "Jobshop like queueing systems," Mgt. Sc., 10, 1, Oct. 1963, 131-142.

Kelly, F.P., "Networks of queues with customers of different types.," Jour. of App. Prob., 12, 1975, 542-554.

Kleinrock, L., Queueing Systems, Vol I, Wiley, Interscience, NY, 1975.

———————— Queueing Systems, Vol II, Wiley, Interscience, NY, 1976.

Kobayashi, H., "Application of diffusion approximations to queueing networks. Part I Equilibrium queue distributions," J. Assoc. Comput. Mach., 21, 1974, 316-328.

Kuehn, P.J., "Analysis of complex queueing networks by decomposition," 8th ITC, Melbourne, 1976, 236.1-236.8.

Kuehn, P.J., "Approximate analysis of general networks by decomposition," IEEE Trans. Comm., COM-27, 1, 1979, 113-126.

Kraemer, W. and M. Langenbach-Belz, "Approximate formulae for the delay in the queueing system GI/G/1," Congressbook, 8th ITC, Melbourne, 1976, 235.1-235.8.

Lemoine, A.J., "Networks of Queues - A survey of equilibrium analysis," Mgt. Sc.,24, 4, Dec. 1977, 464-481.

Newell, G.F., Applications of Queueing Theory, Chapman and Hall, London, 1971, chapter 6.

Reiser, M., and H. Kobayashi, "Accuracy of diffusion approximation for some queueing systems," IBM Jour. of Res. Dev., 18, 1974, 110-124.

Reiser, M. and S.S. Lavenberg, "Mean value analysis of closed multichain networks.," Jour. of ACM, 27, 2, 1980, 313-323.

Schweitzer, P., "Approximate analysis of multiclass closed network of queues," Presented at International Conference, Stochastic control and optimization, 1979, Amsterdam, Netherlands.

Sevick, K.C., A.I. Levy, S.K. Tripathi and J.L. Zahorjan, "Improving approximations of aggregated queueing network systems," Proc. Computer Performance, Modelling, Measurement and Evaluation, 1977.

Shantikumar, J.G., and J.A.Buzacott, "Open queueing network models of dynamic job shops," Int. Jour. of Prod. Res., 19, 3, 1981, 255-266.

Shimshak, D.G., "A comparison of waiting time approximations in series queueing systems," NRLQ, 26, 1979, 499-509.

Suri, R. and M.A. Zazanis, "Perturbation analysis gives strongly consistent estimates for the M/G/1 queue," Sep 1984.

Whitt, W., "Approximating a point process by a renewal process: Two basic methods," Oper. Res., 30, 1, 1982, 125-147.

Whitt, W., "The queueing network analyzer," Bell System Technical Journal, 62, 9, 1983a, 2779-2815.

Whitt, W., "Performance of the queueing network analyzer," Bell System Technical Journal, 62, 9, 1983b, 2817-2843.

Whitt, W., "Approximations for departure processes and queues in series," NRLQ, 31, 1984, 499-521.

Zazanis, M.A., and R. Suri, "Comparison of perturbation analysis with conventional sensitivity estimates for regenerative stochastic systems," Dec 1984.

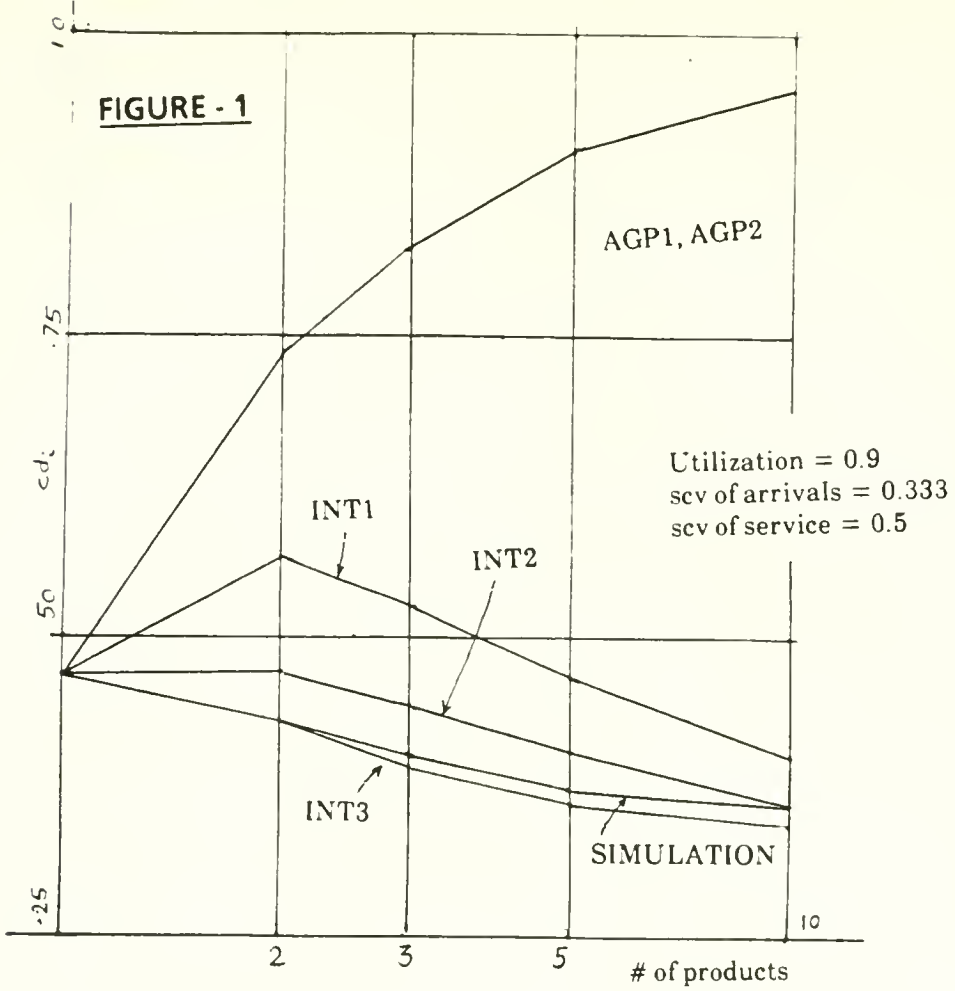


FIGURE - 1
AGP1, AGP2
Utilization = 0.9
scv of arrivals = 0.333
scv of service = 0.5
INT1
INT2
INT3
SIMULATION
cd_i
1.0
.75
.50
.25
2
3
5
10
# of products

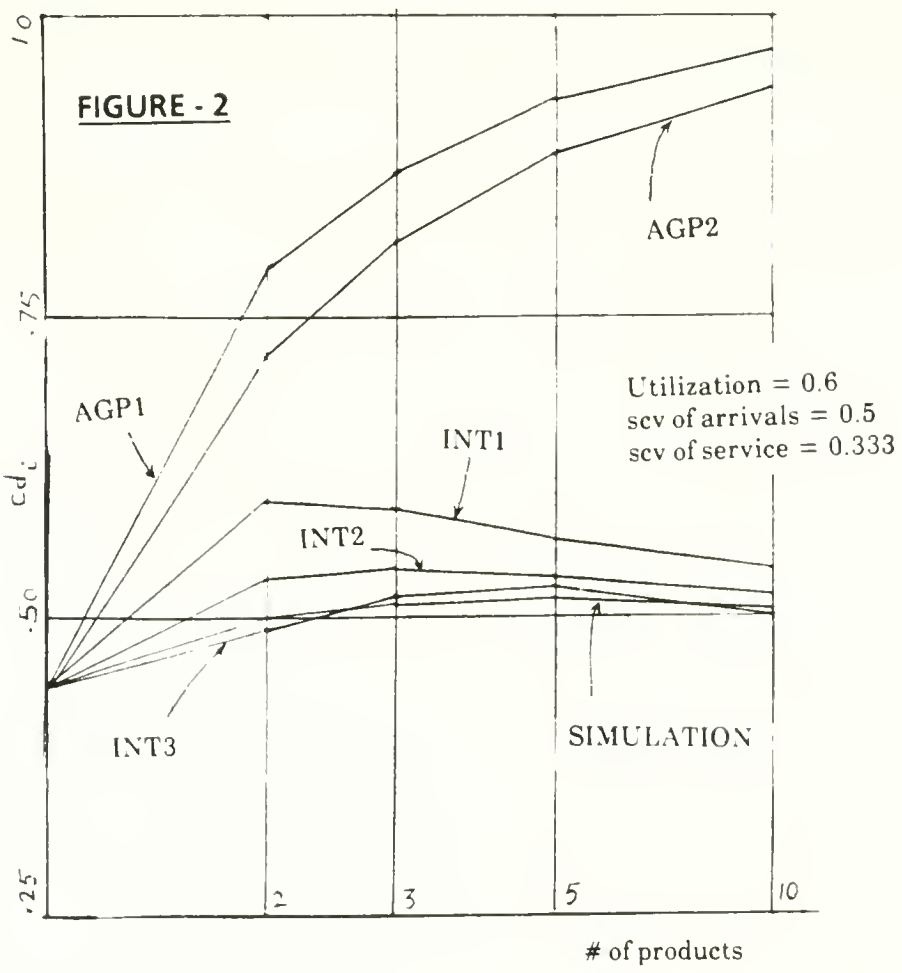
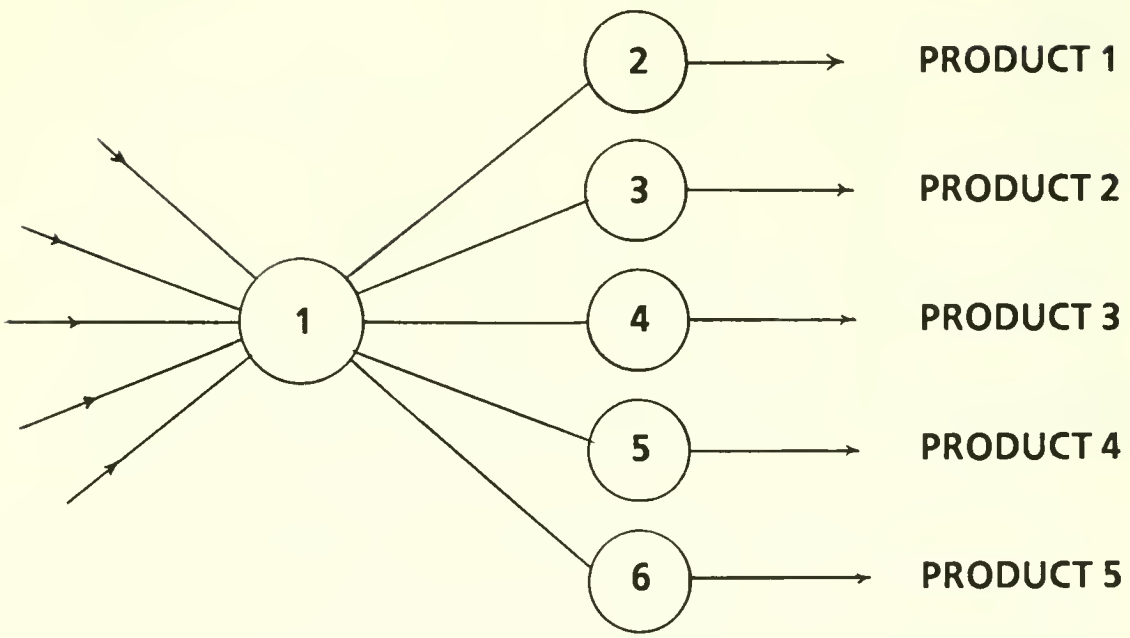


FIGURE - 2
AGP2
AGP1
Utilization = 0.6
scv of arrivals = 0.5
scv of service = 0.333
INT1
INT2
INT3
SIMULATION
cd_i
1.0
.75
.50
.25
2
3
5
10
# of products

Figure 3: The 5 Product, 6 Station Example



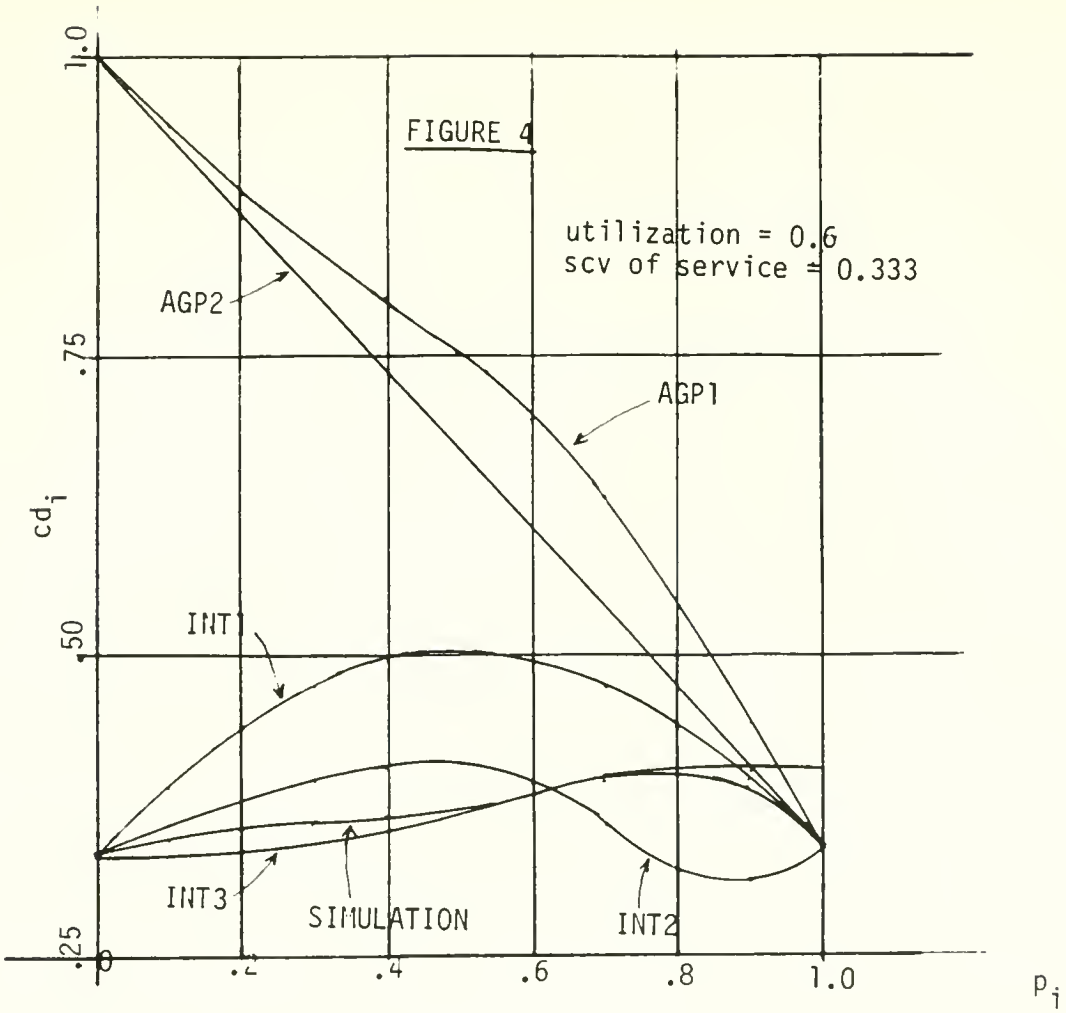


FIGURE 4
utilization = 0.6
scv of service = 0.333
AGP2
AGP1
INT1
INT3
SIMULATION
INT2
1.0
.75
.50
.25
0
.2
.4
.6
.8
1.0
cd_i
p_i

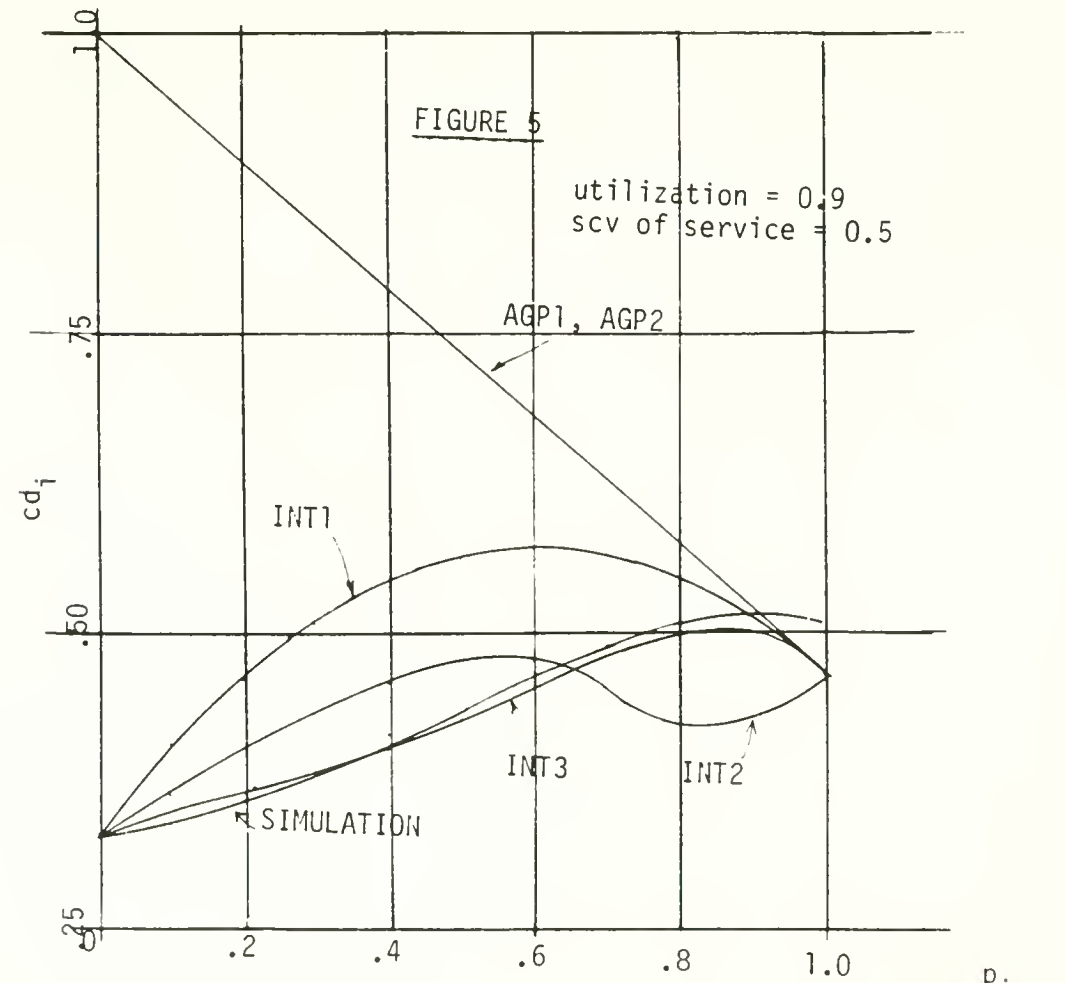


FIGURE 5
utilization = 0.9
scv of service = 0.5
AGP1, AGP2
INT1
INT3
INT2
SIMULATION
1.0
.75
.50
.25
0
.2
.4
.6
.8
1.0
cd_i
p_i

Table 1: Comparison of Approximations - Estimation of cd

Station Utilization $\rho = 0.9$
scv of service = 0.333

# of products	scv of arrivals	AGP1	AGP2	INT1	Simulation
m=2, p=.5	.5	.6842	.6825	.559	.4749
	.333	.669	.6665	.502	.373
	.25	.6614	.6588	.474	.328
m=3 p=.333	.5	.791	.789	.568	.486
	.333	.781	.778	.484	.37
	.25	.776	.773	.442	.3025
m=5 p=.2	.5	.876	.873	.553	.522
	.333	.87	.867	.44	.3606
	.25	.867	.864	.394	.2874
m=10 p=.1	.5	.939	.937	.5315	.5146
	.333	.937	.933	.393	.355
	.25	.936	.932	.324	.2785

Table 2: Comparison of Approximations - Estimation of cd

Station Utilization, $\rho = 0.9$
scv of service = 0.5

# of products	scv of arrivals	AGP1	AGP2	INT1	Simulation
m=2 p=.5	.5	.742	.750	.6172	.5316
	.333	.737	.734	.5698	.4365
	.25	.729	.727	.541	.3921
m=3, p=.333	.5	.836	.834	.613	.5192
	.333	.826	.823	.529	.407
	.25	.821	.818	.487	.355
m=5, p=.2	.5	.903	.9	.58	.52
	.333	.897	.894	.467	.3726
	.25	.894	.891	.41	.3093
m=10, p=.1	.5	.953	.95	.545	.506
	.333	.95	.947	.407	.3612
	.25	.949	.945	.3378	.2875

Table 3: Comparison of Approximations - Estimation of cd

Station Utilization, $\rho = 0.6$
scv of service = 0.3333

# of products	scv of arrivals	AGP1	AGP2	INT1	Simulation
m = 2, p = .5	.5 .333 .25	.782 .75 .734	.72 .667 .64	.595 .499 .453	.5001 .371 .303
m = 3, p = .333	.5 .333 .25	.874 .858 .85	.814 .778 .76	.591 .481 .427	.507 .377 .287
m = 5, p = .2	.5 .333 .25	.934 .928 .925	.888 .867 .856	.568 .44 .376	.51 .36 .2756
m = 10, p = .1	.5 .333 .25	.9713 .9697 .9689	.944 .933 .928	.539 .393 .32	.5074 .354 .2737

Table 4: Comparison of Approximations - Estimation of cd

Station Utilization, $\rho = 0.6$
scv of service = 0.5

# of products	scv of arrivals	AGP1	AGP2	INT1	Simulation
m = 2, p = .5	.5 .333 .25	.813 .78 .764	.75 .697 .67	.625 .53 .482	.54 .396 .333
m = 3, p = .333	.5 .333 .25	.894 .878 .87	.834 .798 .78	.61 .501 .446	.533 .37 .304
m = 5, p = .2	.5 .333 .25	.946 .94 .937	.9 .879 .868	.58 .442 .388	.52 .353 .281
m = 10, p = .1	.5 .333 .25	.977 .976 .975	.95 .939 .934	.545 .399 .326	.512 .36 .275

Table 5: Parameters for the 6 Station, 5 Product Example

	Case 1	Case 2
# of products	5	5
# of stations	6	6
scv of arrivals	0.5	0.333
scv of service	0.5	0.333
utilization - station 1	0.6	0.6
utilization - stations 2-6	0.8	0.8

Table 6: Results for the 6 Station, 5 Product Example

Mean Number of Jobs at Stations 2-6.

	MM1	MG1	AGP1	AGP2	INT1	Sim
Case1	4.0 (68.07)	3.2 (34.45)	3.11 (30.67)	3.04 (27.7)	2.48 (4.2)	2.38 2.37*
Case 2	4.0 (124)	2.93 (64.05)	2.82 (57.9)	2.72 (52.3)	1.96 (9.7)	1.786 1.8*

* is the estimate of L by using the cd obtained from the simulation.
Figures in parenthesis show % deviation from the simulation value

Table 7: Comparison of Approximations - Varying p_i

Station Utilization, $\rho = 0.6$
scv of service = 0.333

p_i	AGP1	AGP2	INT1	INT2	INT3	Simulation		
						Erl	Unif	Beta
0.1	.9386	.933	.393	.354	.335	.3507	.3463	.3410
0.2	.8864	.867	.439	.376	.340	.3521	.336	.358
0.3	.8405	.800	.473	.394	.346	.3618	.3407	.3398
0.4	.7967	.733	.493	.407	.356	.3582	.3412	.3510
0.5	.7499	.6667	.500	.409	.370	.3724	.3433	.3754
0.6	.6951	.600	.493	.392	.383	.3810	.3534	.3885
0.7	.6280	.543	.473	.356	.394	.4009	.3724	.3936
0.8	.5457	.480	.440	.319	.397	.3977	.3603	.3899
0.9	.4473	.400	.393	.310	.383	.4031	.3787	.3981

Table 8: Comparison of Approximations - Varying p_i

Station Utilization, $\rho = 0.6$
scv of service = 0.5

p_i	AGP1	AGP2	INT1	INT2	INT3	Simulation		
						Erl	Unif	Beta
0.1	.9446	.939	.399	.360	.341	.3416	.3346	.3534
0.2	.8984	.879	.452	.388	.352	.3380	.3395	.3597
0.3	.8585	.8179	.490	.4119	.3649	.3794	.3525	.3575
0.4	.8207	.7572	.5172	.4312	.3802	.3900	.3681	.3982
0.5	.7799	.6965	.5265	.4385	.3995	.3964	.3738	.3944
0.6	.7311	.6358	.5288	.4278	.4188	.4154	.3971	.4081
0.7	.6700	.5751	.5151	.3981	.4361	.4372	.4024	.4407
0.8	.5937	.5144	.4874	.3664	.4444	.4637	.4219	.4260
0.9	.5013	.4537	.4467	.3617	.4347	.4646	.4207	.4406

Note:
The results under the subcolumn Erl refer o the estimates of cd obtained by simulation with Erlang distribution with parameter 3. Likewise, Unif and Beta refer to the simulation results with Uniform and Beta distribution respectively.

Table 9: Comparison of Approximations - Varying p_i

Station Utilization, $\rho = 0.9$
scv of service = 0.5

p_i	AGP1	AGP2	INT1	INT2	INT3	Simulation		
						Erl	Unif	Beta
0.1	.9469	.9468	.4068	.3678	.3488	.3389	.3338	.3586
0.2	.8941	.8936	.4666	.4026	.3666	.3604	.3619	.3807
0.3	.8416	.8404	.5134	.4344	.3874	.3911	.3743	.3881
0.4	.7891	.7872	.5472	.4572	.4062	.4215	.4064	.4508
0.5	.7366	.7340	.5670	.4760	.4370	.4387	.4338	.4495
0.6	.6837	.6808	.5738	.4728	.4638	.4923	.5059	.5263
0.7	.6303	.6276	.5676	.4506	.4886	.4923	.5059	.5263
0.8	.5766	.5732	.5462	.4252	.5032	.5074	.5154	.5076
0.9	.5225	.5212	.5142	.4292	.5022	.5143	.5002	.5200

Table 10: Comparison of Approximations - Varying p_i

Station Utilization, $\rho = 0.9$
scv of service = 0.333

p_i	AGP1	AGP2	INT1	INT2	INT3	Simulation		
						Erl	Unif	Beta
0.1	.9334	.9330	.393	.354	.335	.3300	.3234	.3421
0.2	.8671	.8670	.439	.376	.340	.3385	.3308	.3447
0.3	.8011	.8000	.473	.394	.347	.3554	.3552	.3437
0.4	.7351	.733	.493	.407	.356	.3794	.3683	.3588
0.5	.6691	.667	.500	.409	.370	.3739	.3660	.3848
0.6	.6027	.600	.493	.392	.383	.3945	.3858	.3958
0.7	.5358	.5331	.473	.356	.394	.4035	.4118	.3937
0.8	.4686	.4664	.440	.319	.397	.4088	.3987	.4054
0.9	.4010	.400	.393	.310	.383	.3860	.3809	.4020

Note:
The results under the subcolumn Erl refer o the estimates of cd obtained by simulation with Erlang distribution with parameter 3. Likewise, Unif and Beta refer to the simulation results with Uniform and Beta distribution respectively.

Table 11: Performance of Approximations INT2 and INT3 with Multiple Products

Estimation of cd, the scv of departure streams

# of products	ca_1	util = .9 cs = .333			util = .9 cs = 0.5			util = .6 cs = .333			util = .6 cs = 0.5		
		INT2	INT3	Sim	INT2	INT3	Sim	INT2	INT3	Sim	INT2	INT3	Sim
2	0.5	.4945	4603	4749	.5620	5278	.5316	.5320	4978	.5001	.5620	5278	5400
	0.333	4080	.3690	3725	.4760	4370	4365	.4080	.3690	.3710	4365	3995	3956
	0.25	.3624	.3227	3275	.4305	.3908	.3921	.3440	.3043	3029	.3740	3343	.3334
3	0.5	.5170	.4692	4856	.5620	.5142	.5192	.5410	.4932	5069	.5620	.5142	5330
	0.333	.3990	.3507	.3697	4440	3957	.4070	3990	.3507	.3765	.4190	.3707	.3704
	0.25	.3338	.2902	.3025	.3790	.3354	.3549	.3210	.2774	.2868	.3390	.2954	.3035
5	0.5	.5350	.4959	.5220	.5690	.5299	.5203	.5320	.4929	.5100	.5690	.5299	5200
	0.333	.3756	.3396	3606	.4026	.3666	.3726	3756	3396	.3600	.3876	.3516	.3525
	0 25	.3005	.2698	2874	.3276	2969	.3093	.2930	.2623	2756	.3050	.2743	2810
10	0.5	.5080	4880	.5146	.5190	.5010	.5062	.5130	.4950	.5074	.5190	.5010	.5115
	0.333	.3540	.3345	.3546	.3678	.3483	.3612	.3540	.3345	.3540	.3603	.3408	.3600
	0 25	.2760	.2565	2785	.2893	.2718	.2875	.2720	.2545	2737	2760	.2605	2753

Table 12: Performance of INT2 and INT3 for the 6 Station, 5 Product Example

Estimate of Mean number of jobs at Stations 2-6.

	INT2	INT3	Simulation
Case 1 (% Error)	2.416 (1.51)	2.344 (1.51)	2.38
Case 2 (% Error)	1.818 (2.13)	1.766 (1.12)	1.786

Table 15: Summary of Recommendations - Choice of Approximations

	$ca \leq 0.5$	$ca > 0.5$
$p_i \leq 0.5$	INT2 or INT3	INT1
$p_i > 0.5$	INT3	INT1

Table 13: Comparison of the Approximations for the Network Experiment

L = Mean number of jobs in the network.

e = % absolute error in the estimate of L relative to the simulation value.

$\bar{e}$ is the average of e for the ten problems.

Case #	Sim. L	INT2		INT3		INT1		AGP2		AGP1	
		L	e	L	e	L	e	L	e	L	e
1	33.4662	32.0013	4.36	31.7464	5.13	36.5375	9.18	43.6134	30.3	45.2174	35.1
2	38.6601	38.4733	0.48	37.9593	1.81	42.0800	8.85	50.8733	31.6	52.3760	35.5
3	32.6694	32.7630	0.29	32.4088	0.80	35.7614	9.46	41.9501	28.4	43.1830	32.2
4	31.6133	32.1740	1.77	31.5609	0.17	34.7732	10.0	41.2094	30.4	42.6819	34.8
5	35.7684	36.3285	1.56	36.0532	0.80	37.2631	4.18	42.8179	19.7	44.1258	23.4
6	36.8333	36.8923	0.16	35.9443	2.41	38.5374	4.62	45.3844	23.2	46.6600	26.7
7	48.0490	47.3808	1.39	46.6743	2.86	50.1025	4.27	58.1530	21.0	59.5882	24.0
8	47.1312	46.1364	2.11	45.4935	3.47	49.2480	4.49	56.2997	19.5	58.1702	23.4
9	34.7903	33.2916	4.31	33.0425	5.02	35.8511	3.05	41.7618	20.1	42.9693	23.5
10	46.1585	44.4197	3.76	44.2491	4.14	46.8115	1.41	51.9585	12.6	53.2057	15.3
$\bar{e}$			1.78		2.66		5.95		23.7		27.4

Table 14: Network Experiment - Comparison at Station Level

The table provides the distribution of station level errors in the estimates of L for each approximation. The figures represent the number of stations in each category as a % of the total number of observations (130).

Range of error e, at the station level	INT2	INT3	INT1	AGP2	AGP1
$0 \leqq e < 5.0$	56.16	49.23	40.77	7.69	0.77
$5.0 \leqq e < 10.0$	30.00	32.31	30.77	3.08	6.15
$10.0 \leqq e < 15.0$	7.69	14.61	14.62	12.31	8.47
$15.0 \leqq e < 20.0$	6.15	3.85	7.69	17.69	17.69
$20.0 \leqq e$	0.00	0.00	6.15	59.23	66.92
Max. error	18.8	18.8	30.62	63.37	64.39

APPENDIX: Data Base for the Network Example

Number of Stations = 13.

Number of products = 10

Product Routing Characteristics:

Product	number of operations	Routing sequence*
1	7	1,2,4,2,9,10,11
2	8	1,2,5,2,8,9,10,11
3	8	1,2,6,4,2,9,12,11
4	8	1,2,7,4,2,9,10,11
5	8	1,2,4,12,2,9,2,13
6	8	1,2,5,12,2,9,7,13
7	8	1,2,6,12,2,8,2,13
8	12	1,2,3,7,4,12,2,8,6,9,2,13
9	13	1,2,3,5,4,6,12,2,8,2,10,6,13
10	13	1,2,3,6,2,4,12,7,2,9,11,5,13

* The routing sequence for each product specifies the stations visited by the product in order.

9911 008

Date Due BASEMENT

OCT.

FEB 1 5 1997

JUN 0 8 1997

JUN 0 8 1997

JUL 0 8 1997

FEB

JUL 2 6 1998

SEP. 3 0 1998

Lib-26-67

3 9080 004 092 307